



Article

# Characteristics and Comparative Genomic Analysis of a Novel Virus, VarioGold, the First Bacteriophage of *Variovorax*

Przemyslaw Decewicz , Michal Kitowicz and Monika Radlinska \*

Department of Environmental Microbiology and Biotechnology, Institute of Microbiology, Faculty of Biology, University of Warsaw, Miecznikowa 1, 02-096 Warsaw, Poland

\* Correspondence: m.radlinska@uw.edu.pl; Tel.: +48-22-5541411

**Abstract:** *Variovorax* represents a widespread and ecologically significant genus of soil bacteria. Despite the ecological importance of these bacteria, our knowledge about the viruses infecting *Variovorax* spp. is quite poor. This study describes the isolation and characterization of the mitomycin-induced phage, named VarioGold. To the best of our knowledge, VarioGold represents the first characterized virus for this genus. Comparative genomic analyses suggested that VarioGold is distinct from currently known bacteriophages at both the nucleotide and protein levels; thus, it could be considered a new virus genus. In addition, another 37 prophages were distinguished in silico within the complete genomic sequences of *Variovorax* spp. that are available in public databases. The similarity networking analysis highlighted their general high diversity, which, despite clustering with previously described phages, shows their unique genetic load. Therefore, the novelty of *Variovorax* phages warrants the great enrichment of databases, which could, in turn, improve bioinformatic strategies for finding (pro)phages.

**Keywords:** *Variovorax*; bacteriophage; prophage; comparative genomic; Zloty Stok



**Citation:** Decewicz, P.; Kitowicz, M.; Radlinska, M. Characteristics and Comparative Genomic Analysis of a Novel Virus, VarioGold, the First Bacteriophage of *Variovorax*. *Int. J. Mol. Sci.* **2022**, *23*, 13539. <https://doi.org/10.3390/ijms232113539>

Academic Editor: Alicja Węgrzyn

Received: 6 October 2022

Accepted: 31 October 2022

Published: 4 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

*Variovorax* are Gram-negative, aerobic and motile bacteria belonging to the family *Comamonadaceae* of the *Betaproteobacteria* class. Members of this genus are commonly present in soil and freshwater environments and have been isolated from many locations in Europe, Asia and the Americas, as well as from polar regions [1]. Interestingly, *V. paradoxus* has also been identified as part of the methylotrophic microbiota of the human mouth [2]. The diverse distribution of these bacteria is in line with their ability to grow over a wide temperature range, as meso- and psychrophilic strains have also been discovered [3]. *Variovorax* spp. are metabolically diverse and exhibit the ability to degrade a variety of substrates [1], including a series of organic pollutants, such as BTEX (benzene, toluene, ethylbenzene and xylene), phenol, trichloroethylene (TCE), acrylamide and pesticides [4–7]. Furthermore, there has also been a report of a strain that is capable of the biotransformation of ibuprofen, using it as a sole carbon and energy source [8,9]. Another interesting feature is the ability to degrade acyl-homoserine lactones, which are signaling molecules used by many *Proteobacteria* for social communication in a phenomenon known as ‘quorum sensing’ [9].

The *Variovorax* genus is also a member of the group of plant growth-promoting rhizobacteria (PGPR), having a direct or indirect influence on the host plant; thus, it is used as model bacteria for the study of microbe–plant interactions [10]. It was reported that, in some cases, the presence of a rhizobacterium enhances the plant’s resistance to biotic and abiotic stresses [11]. Thus far, *Variovorax* spp. strains isolated from the rhizospheres of important crops, such as sunflower (*Helianthus annuus* L.) or tomato (*Solanum lycopersicum* L.), have been described [12,13]. They were also isolated from the leaf surface of the common dandelion (*Taraxacum officinale* F.H. Wigg.) and from lichen *Himantormia* [14].

Several studies have reported the extreme heavy metal tolerance of the *Variovorax* representatives that originate from metal-polluted habitats (industrialized and post-industrial areas, flotation tailing dumps, mining waste). For example, the cadmium- and cobalt-resistant 5C-2 strain was isolated from the root zone of Indian mustard (*Brassica juncea* L.) cultivated on mining waste [15,16], and RA128A was found to be resistant to multiple heavy metals, including Zn, Pb, Cd, Cu and Ag [17]. In turn, IDSBO-4 is capable of arsenite [As(III)] and antimonite [Sb(III)] oxidation [18]. The unique properties of *Variovorax* spp. make them promising candidates for research into the bioremediation of soils contaminated with heavy metal compounds [19,20].

In this work, we describe a novel *Variovorax* sp. strain, ZS18.2.2, a biofilm inhabitant in the Zloty Stok gold and arsenic mine (Poland). Unlike the strains mentioned above, which originate from the soil of the mine area or mine tailings, ZS18.2.2 was isolated from a sample of a natural microbial biofilm that covered the rock roof and walls located at the end of the two-kilometer-long Gertruda Adit in the mine [21]. The abandoned Zloty Stok mine is characterized by harsh environmental conditions in terms of a stable low temperature (~10 °C), reduced concentration of oxygen (17.2%) and very high concentrations of metals, including an abundance of arsenic [21,22]. A number of arsenic-resistant strains from this site have already been characterized [23–25]. Interestingly, all of them were lysogens—as was the ZS18.2.2.

Bacterial viruses (bacteriophages; phages in short) are the most abundant biological entities on Earth, with an estimated  $10^{31}$  virus-like particles in the biosphere [26]. They play a significant role in the life cycle and evolution of each bacterial genus and maintain microbial balancing in each ecosystem [27]. Most known phages have double-stranded DNA genomes packed into a tailed proteinaceous capsid. They can either be virulent or temperate [28]. Virulent phages only reproduce via the lytic cycle, while temperate phages can undergo either a lytic or a lysogenic cycle. In the latter, viral DNA persists in the host as a prophage that is integrated into the bacterial host chromosome or maintained extrachromosomally in the cytoplasm as independent episomes, replicating alongside the host DNA [29]. Prophages are a major reservoir of new genes for bacteria and can provide multiple benefits to the host—for example, in niche adaptation, biofilm formation and the production of virulence factors; they increase the host's response to general environmental stresses and increase the host's resistance to antibiotics and superinfection [30–32]. Understanding viral contributions to the evolution of bacteria and the possibility for the reshaping of bacterial cell functions by infection are some of the reasons for the renewed interest in bacteriophages. In addition, phages are commonly used to develop new genetic tools for routine genetic manipulation, such as replicative vectors and phage delivery systems [33]. Furthermore, phage-encoded proteins (e.g., depolymerases) seem to be promising tools for controlling pathogenic bacteria and other applications, including clinical diagnostics and biochemical analyses [34].

Although the number of sequenced phage genomes continues to increase, viruses infecting many ecologically important and abundant bacteria still remain poorly investigated. For example, there is a large disproportion in the number of characterized phages infecting highly studied *Gammaproteobacteria* and (pro)phages of *Betaproteobacteria* [35], which in turn hampers the understanding of the ecological roles and biological traits of the latter viruses. Among the group of understudied phages are those infecting the *Variovorax* genus. We found only one report on the lysogenicity of various soil isolates, which resulted in the identification of *Variovorax* hosts containing inducible prophages [36]; however, this mentioned survey was limited to only morphological studies of isolated viruses. In contrast to this, our study describes not only the isolation, but also contains a thorough analysis of the architecture of the *Variovorax* temperate bacteriophage genome, as well as its comparative analysis with known viruses and with putative prophage sequences that were identified by us in complete *Variovorax* genomes. Therefore, this is the first report of an active *Variovorax* phage that includes its complete genome sequence and characterization. These findings will help to fill the gap in our understanding of previously unrecognized

*Variovorax* phages and substantially expand our current knowledge regarding the genomic diversity of bacterial viruses.

## 2. Results and Discussion

### 2.1. Identification and Characterization of *Variovorax* sp. strain ZS18.2.2

A new *Variovorax* sp. strain, designated ZS18.2.2, was isolated from a sample of rock biofilms from the Zloty Stok gold and arsenic mine—collected in February 2021. When grown on solid R2A medium, this strain formed colonies with mucoid morphology (Figure S1), possibly due to its EPS-producing ability, which suggests that the strain is actively involved in biofilm production; moreover, surface motility, similar to swarming motion, has also been observed at 4, 10 and 20 °C on semi-solid (0.5%) R2A and LB media, and at 30 °C only on R2A. Similar properties were reported and examined for the reference strain, *V. paradoxus* EPS [13]. To determine the temperature requirements of *Variovorax* sp. ZS18.2.2 for growth, we performed a temperature tolerance analysis on plates with LB and R2A media. This test revealed that ZS18.2.2 grew in both media at temperatures ranging between 4 °C and 30 °C (but not at 37 °C). Therefore, this bacterium can be considered psychrotolerant and, consequently, well adapted to living at the stable low temperatures of the Zloty Stok gold mine.

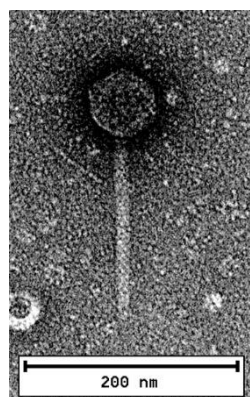
As ZS18.2.2 was isolated from an arsenic-rich environment where the concentration of arsenic hydride reached 1.52–3.23 mg/m<sup>3</sup> [21], the resistance of this strain to inorganic arsenic species was tested. ZS18.2.2 showed extreme tolerance to As(V) (up to 300 mM) and moderate resistance to As(III) (up to 10 mM). Therefore, *Variovorax* sp. ZS18.2.2 is another strain isolated from the Zloty Stok mine that has shown considerable arsenic resistance; moreover, when ZS18.2.2 grows in R2A medium containing 5 or 10 mM of As(III), it oxidizes As(III) to As(V). In comparison to other known *Variovorax* strains that are resistant to As(III) or As(V), ZS18.2.2 was found to be much more tolerant to As(V) than e.g., *Variovorax* sp. MM-1 (200 mM) [37], but less tolerant to As(III) than *V. paradoxus* Dhal F [38] and *Variovorax* sp. MM-1 [37], which showed resistance up to 15 mM and 20 mM, respectively. To the best of our knowledge, *Variovorax* sp. ZS18.2.2 shows the highest resistance to As(V) compounds among all *Variovorax* strains that have been isolated and tested for this characteristic so far (Table S1).

The total DNA of ZS18.2.2 was isolated and sequenced. The reconstruction of its genome resulted in 46 contigs with a total length of 7,235,762 bp and 66.32% average GC content. A putative extrachromosomal element—plasmid (112,571 bp)—was found (JANLNM010000015).

We searched the ZS18.2.2 genome for putative prophage sequences using the PhiSpy program [39]. The analysis revealed the presence of such a region, within contig 3 (JANLNM010000003). To verify whether the recognized prophage region was an active virus, we treated the ZS18.2.2 cells with mitomycin C, a classical inducer of lambdoid prophages. This approach caused the induction of the phage named VarioGold. Its DNA was isolated from capsids and re-sequenced. The obtained reads confirmed the previously recovered prophage sequence and also allowed for the determination of phage termini (see below).

### 2.2. Morphological Analysis by TEM

The VarioGold phage was subjected to TEM analysis to determine its morphotype. Its virions had an icosahedral head (60 nm in diameter) and a flexible, non-contractile tail (approximately 170 nm long), thus showing a siphovirus morphotype (Figure 1).



**Figure 1.** Visualization of the VarioGold virion.

### 2.3. Host Range Testing

The ZS18.2.2 strain was the only member of the *Variovorax* genus isolated from the Zloty Stok gold mine. Therefore, we used two model strains of *V. paradoxus* EPS and B4 as potential hosts for VarioGold in the spot test. None of the tested strains supported the detectable lytic growth of VarioGold. Presumably, VarioGold is highly specific and has a narrow host range, possibly confined to its natural host strain, ZS18.2.2—although this hypothesis needs experimental validation with the use of a larger number of *Variovorax* strains.

### 2.4. Genomic Analysis of the VarioGold Phage

#### 2.4.1. Identification of the Genome Termini of the VarioGold Phage

The genome of the VarioGold phage consists of 39,429 bp linear double-stranded DNA molecules with a GC content of 62.7%, which is somewhat lower than that of the host (66.32%; see above). The VarioGold genome termini and its possible packaging strategy were analyzed using the PhageTerm tool [40]. This program predicted that VarioGold had linear genomic DNA with fixed termini, i.e., 10 nt 3' cohesive ends, and that the protruding overhang had a nucleotide sequence of CGATCGGTTC. This suggests that VarioGold utilizes the cohesive-end packaging strategy to package its genome. As cohesive ends are covalently joined together in the prophage state, forming a so-called cos site, we compared the VarioGold prophage sequence within contig 3 of the ZS18.2.2 draft genome with those of the corresponding regions of the DNA isolated from free phage particles. This alignment revealed that the VarioGold prophage indeed contains one 10 bp cos site.

#### 2.4.2. Identification of the VarioGold Attachment Site

Most integrases of prokaryotic viruses use the tRNA and tmRNA genes of its host as integration sites [41]. We discovered that the VarioGold prophage is flanked by duplicates of the 15 bp sequence (CCTCCCTCTCCTCCA), which most likely constitutes the bacterial attachment site (attB). One of its copies forms the 3' part of a tRNA-Ser (CGA)-encoding gene that is located 176 nucleotides upstream of the 5' end of the putative integrase gene (VG\_p29). There is only one copy of this 15 bp sequence in the VarioGold genomic DNA that was isolated from capsids.

#### 2.4.3. Module Analysis of the VarioGold Genome

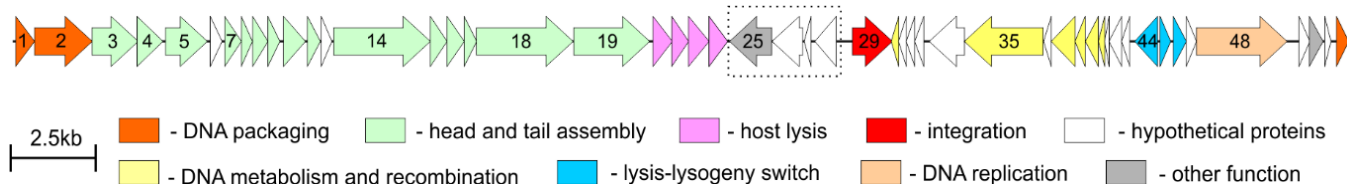
The VarioGold genome was predicted to contain 52 open reading frames (ORFs), with 33 genes (63%) located on the sense strand and 19 genes (37%) on the antisense strand. A function was assigned to 38 ORFs (73% of the total gene number). The remaining 14 ORF-encoded proteins showed similarities with hypothetical proteins that were already described but were of an unknown function. No gene-encoding tRNA was detected in the genome. The positions, sizes and putative functions of the proteins are listed in Table 1.

**Table 1.** Genes located within the VarioGold phage.

| ORF | Coding Region (bp) | Strand | Protein Size (aa) | Predicted Function  |
|-----|--------------------|--------|-------------------|---|
| 1   | 99..653            | +      | 184               | Terminase small subunit   |
| 2   | 653..2350          | +      | 565               | Terminase large subunit   |
| 3   | 2347..3702         | +      | 451               | Portal protein  |
| 4   | 3668..4429         | +      | 253               | Protease/scaffold protein   |
| 5   | 4506..5735         | +      | 409               | Major capsid protein  |
| 6   | 5821..6168         | +      | 115               | Hypothetical protein  |
| 7   | 6245..6736         | +      | 163               | Head-tail connector protein   |
| 8   | 6733..7059         | +      | 108               | Head closure protein  |
| 9   | 7067..7516         | +      | 149               | Tail-component  |
| 10  | 7513..7866         | +      | 117               | Tail completion protein   |
| 11  | 7974..8618         | +      | 214               | Tail protein  |
| 12  | 8677..9057         | +      | 126               | Tail assembly chaperone   |
| 13  | 9084..9395         | +      | 103               | DUF1799 domain-containing phage protein                               |
| 14  | 9446..12,292       | +      | 948               | Tail length tape-measure protein H                                    |
| 15  | 12,292..12,783     | +      | 163               | Minor tail protein  |
| 16  | 12,783..13,289     | +      | 168               | Minor tail protein  |
| 17  | 13,286..13,654     | +      | 122               | NlpC/P60 family protein   |
| 18  | 13,651..16,491     | +      | 946               | Tip attachment protein J  |
| 19  | 16,511..18,742     | +      | 743               | Tail spike-like protein   |
| 20  | 18,841..19,407     | +      | 188               | 4 TM segments containing protein                                      |
| 21  | 19,391..19,888     | +      | 165               | 3 TM segments containing protein; holin                               |
| 22  | 19,885..20,490     | +      | 201               | Phage lysozyme  |
| 23  | 20,487..21,035     | +      | 182               | Spanin, inner membrane subunit  |
| 24  | 20,731..20,997     | +      | 88                | Spanin, outer lipoprotein subunit                                     |
| 25  | 21,090..22,325     | —      | 411               | 10 TM segments containing protein;<br>O-antigen ligase family protein |
| 26  | 22,340..23,239     | —      | 299               | Hypothetical protein; N-term signal peptide                           |
| 27  | 23,286..23,486     | —      | 66                | 2 TM segments containing protein                                      |
| 28  | 23,599..24,228     | —      | 209               | 2 Transmembrane segments containing protein                           |
| 29  | 24,708..25,880     | +      | 390               | Tyrosine recombinase/integrase  |
| 30  | 26,063..25,845     | —      | 72                | MerR family regulatory protein  |
| 31  | 26,060..26,290     | —      | 76                | Hypothetical protein  |
| 32  | 26,287..26,538     | —      | 83                | Hypothetical protein  |
| 33  | 26,535..26,804     | —      | 89                | Hypothetical protein  |
| 34  | 26,963..27,997     | —      | 344               | Ead/Ea22-like protein   |
| 35  | 27,994..30,309     | —      | 771               | ParB partition protein family   |
| 36  | 30,325..30,552     | —      | 75                | Hypothetical protein  |
| 37  | 30,549..31,253     | —      | 234               | Deoxynucleoside monophosphate kinase                                  |
| 38  | 31,250..31,546     | —      | 98                | AAA+-type ATPase  |
| 39  | 31,556..31,936     | —      | 126               | ssDNA-binding protein   |
| 40  | 31,936..32,121     | —      | 61                | TM segment containing protein   |
| 41  | 32,118..32,252     | —      | 44                | Hypothetical protein  |
| 42  | 32,249..32,623     | —      | 124               | Hypothetical protein  |
| 43  | 32,620..32,859     | —      | 79                | Hypothetical protein  |
| 44  | 33,006..33,674     | —      | 222               | Repressor protein CI, S24 family peptidase                            |
| 45  | 33,751..34,062     | +      | 103               | DNA-binding transcriptional regulator Cro-like                        |
| 46  | 34,140..34,523     | +      | 127               | CII-like protein, XRE-type HTH domain                                 |
| 47  | 34,520..34,822     | +      | 100               | Hypothetical protein  |
| 48  | 34,822..37,485     | +      | 887               | Toprim domain containing protein                                      |
| 49  | 37,836..38,153     | +      | 105               | Hypothetical protein  |
| 50  | 38,140..38,544     | +      | 134               | Antiterminator Q protein  |
| 51  | 38,591..38,776     | +      | 61                | Hypothetical protein  |
| 52  | 38,939..39,298     | +      | 119               | HNH endonuclease  |



VarioGold displays a modular genome that can be divided into five functional modules: DNA packaging; virion structure and morphogenesis; host lysis; integration–excision; lysis–lysogeny switch; and genome replication (Figure 2).



**Figure 2.** Genome map and functional annotation of the predicted ORFs of the VarioGold phage. Each arrow represents an ORF and its direction corresponds to the direction of gene transcription. The numbers on arrows refer to the ORF number in the genome, e.g., 1—VG\_p01. The white color of arrows refers to genes encoding proteins with unknown function. The dotted line indicates a cluster of genes (including VG\_P25 encoding O-antigen ligase) that are predicted to be jointly regulated and expressed in the VarioGold prophage (putative moron-like genes).

In all members of the *Caudoviricetes* class, a complex of two proteins commonly referred to as terminases accomplish DNA encapsidation when preformed empty procapsids are subsequently filled with the viral genome by means of a DNA packaging machine [42]. The protein products of the VG\_p01 and VG\_p02 genes that are probably involved in the VarioGold genome packaging showed the highest sequence similarity to their counterparts encoded by myovirus *Burkholderia* BgVeeder33 (UEW68542.1, UEW68541.1, 31% and 54% amino acid identity, respectively) and siphovirus *Marinobacter* AS1 (MK088078.1, MK088078.1, 29% and 43% identity, respectively). In many sequenced tail phage genomes, a gene-encoding HNH endonuclease is located next to their cohesive end site and terminase genes, suggesting a role of HNH proteins in the endonuclease and/or packaging activities of the terminases [43]. Therefore, the position in the VarioGold prophage of the VG\_p52 gene (whose protein product contains the putative HNH endonuclease domain, PF01844)—immediately adjacent to the VG\_p01 and VG\_p02 genes—suggests its potential involvement in DNA packaging. Interestingly, HNH endonucleases encoded by genes adjacent to the terminases of the abovementioned BgVeeder33 and AS1 phages showed similarity to VG\_p52 (38 and 35% identity, respectively).

Eighteen ORFs located downstream of the packaging module may have assigned functions based on their homology to structural proteins encoded by other phages: portal (p03), scaffold (p04), major head (p05), head–tail connector (p08), head closure tail component (p09) and tail (p10–p19) proteins. In the case of VG\_p06, no viral counterpart has been found. This protein had little similarity with a few hypothetical *Betaproteobacteria* proteins. Interestingly, the HHpred search for a putative tail spike of VarioGold (VG\_p19) identified a hit (with 100% probability, E value,  $3.4 \times 10^{-45}$ ) with pectate lyase superfamily proteins. These types of virion-associated enzymes are used by phages for the degradation of bacterial capsular polysaccharides and are called depolymerases [34,44].

The virion structural module in the VarioGold genome is followed by a cluster of five genes (VG\_p20–p24) that were predicted to comprise the lysis module. The VG\_p22 protein contains a lysozyme-like motif (pfam00959); thus, it presumably acts as an endolysin (murein hydrolase) that cleaves the  $\beta$ -1,4-linkages between adjacent N-acetylmuramic acid and N-acetylglucosamine residues in cell wall peptidoglycans. The gene located upstream of VG\_p22 encodes a putative holin (VG\_p21) and was predicted to contain three transmembrane helices and an N-terminal signal peptide. VG\_p23 and VG\_p24 are probably the inner and outer membrane subunits of the spanin complex; they have a single N-terminal transmembrane domain and a lipoprotein outer membrane localization signal, respectively. A BLASTP search in the NCBI virus database revealed a homolog for VG\_p22 only (i.e., YP\_009100017.1 of *Escherichia* phage vB\_EcoM-ep3, 39% identity).

The VG\_p29 gene is predicted to encode an integrase, as its protein product belongs to the tyrosine recombinase family (pfamPF00589). Its most closely related viral proteins are tyrosine integrases from *Rhodobacter* siphophages, e.g., RcPutin (GenBank accession QXN72045.1), and RcRios (QXN72145.1), which share 39% identity with the VG\_p29 protein. The putative VarioGold integrase gene is located in the opposite orientation to the segment of 15 mostly hypothetical genes (unassigned function). The functional domains were detected in VG\_p35 (ParB-like nuclease domain, pfam02195), VG\_p37 (Nucleoside/nucleotide kinase superfamily, cl17190), VG\_p38 (P-loop ATPase family, cl38936) and VG\_p39, which was predicted as a putative single-stranded DNA-binding protein.

The last gene (towards the 3' genome end) of this segment is VG\_p44, which, together with the oppositely oriented VG\_p45, comprises a putative lysis-lysogeny locus that is responsible for switching between the lytic and the lysogenic cycles. The carboxyl terminal region of VG\_p44 contains the S24 signal peptidase domain (pfam00717) that is typical of CI repressor-like proteins. Its homologs were found among protein products encoded in several assembled genomes of uncultured viruses and *Burkholderia* phage vB\_BmuP\_KL4 (YP\_009800704.1, 35% identity). The putative transcriptional regulators, Cro-like VG\_p45 and CII-like VG\_p46 (PF06892), had no homologs in the NCBI viral database but showed similarity with those encoded in bacterial genomes.

Two domains were identified in VG\_p48, including Toprim (PF13362; from 227 to 302), which is found in bacterial DnaG-type primases and their viral counterparts [45], and virulence-associated protein E (PF05272; from 574 to 755). While the segment of VG\_p48 spanning the latter domain showed similarity to several putative phage primases (e.g., *Marinobacter* phage PS3, ATN93361.1 [46]), VG\_p48 was homologous over its entire length only with the protein (DAS38011.1) encoded by a virus assembled from human metagenomes [47].

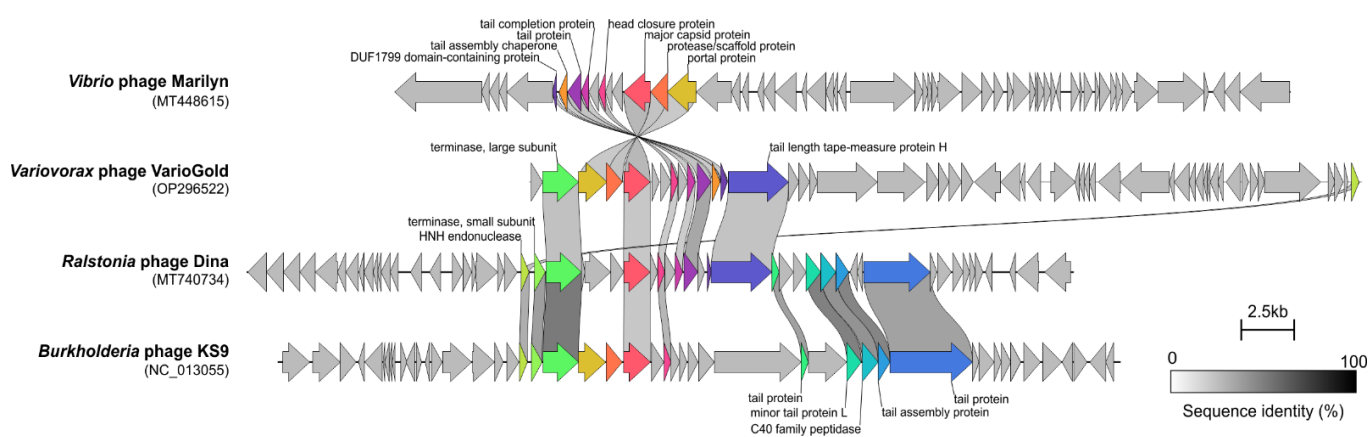
A putative protein product of the VG\_p50 gene showed no viral homologs, but the HHpred search identified a hit for antiterminator Q-like proteins (PF06530) that modify the RNA polymerase near the phage late-gene promoters and thereby cause antitermination at distant sites [48].

The most intriguing gene of VarioGold is VG\_p25, whose protein product shows a 31% amino acid identity with the O-antigen ligases encoded by *Pseudomonas* phages D3 (NP\_061525.2) [49], phi297 (YP\_005098089.1) [50] and PP9W (UAW06749.1). A conserved Wzy\_C superfamily motif (also called an O-Ag ligase domain due to its prevalence in WaaL proteins found in various Gram-negative species that catalyze a key step in lipopolysaccharide synthesis [51]) was identified within the C-terminal half of the VG\_p25 protein (residues 200–326); moreover, the TMPred and TMHMM programs predicted ten strong transmembrane regions in VG\_p25. The presence of multiple transmembrane helices is a characteristic feature of the O-Ag ligase family proteins [52]. It was revealed that the D3 phage causes a serotyping switch of *P. aeruginosa* serotype O5 to O16; thus, during lysogeny, the D3-modified LPS receptors of its host are resistant to LPS-dependent phages. Three proteins of the D3 phage are responsible for this seroconversion phenomenon: an O-acetylase (Oac, NP\_061524.1), the abovementioned O-Ag ligase (NP\_061525.2), and an  $\alpha$ -polymerase inhibitor (Iap, YP\_009173780.1). The genes that encode these proteins are not organized as an operon, as *iap* and O-Ag ligase are located on the complementary strand to *oac*, which is located on the top strand between these genes [53]. The VG\_p25 gene is adjacent to the oppositely oriented host cell lysis gene cluster (VG\_p20-p24; see above) and to the similarly oriented three hypothetical genes (VG\_p26-p28). The common feature of protein products of the latter is the presence of transmembrane helices and/or an N-terminal signal sequence; however, none of these VarioGold putative proteins are similar to the Iap and Oac encoded by the D3 phage. Although it is highly speculative, we suppose that the putative O-antigen ligase gene of VarioGold and the other upstream-located genes (VG\_p26-p28) form a common, collectively regulated cluster because they are all oriented in the same direction, unlike the late genes of this phage (that is, those of replication, structural and host lysis modules); if so, perhaps they are active in lysogeny.

## 2.5. Comparative Genomic Analyses

### 2.5.1. Comparative Analysis of VarioGold with Other Phages

A nucleotide BLAST search of the NCBI viral genome database using the complete genome sequence of VarioGold did not reveal any significant matches covering more than 3% of the query genome. Therefore, to determine the location of VarioGold within the phage population network, a viral cluster analysis with vConTACT2 [54] was conducted, and the ViPTree whole-genome-based phylogenomic tree [55] was constructed. The network indicated that VarioGold was an outlier, and it was exclusively connected via edges with three temperate siphoviruses—two infecting members of *Alphaproteobacteria* (genera: *Ralstonia* and *Burkholderia*) and one infecting the *Vibrio* genus of *Gammaproteobacteria*. The Clinker alignment showed that VarioGold shared eight, seven or six similar DNA packaging and structural proteins with them (Figure 3); these were: (i) *Vibrio* phage Marilyn (portal, protease, major capsid, head closure, tail completion protein, tail protein, tail assembly chaperone, DUF1799 domain-containing phage protein); (ii) *Ralstonia* phage Dina (HNH, TerL, major capsid, head closure, tail completion, tail and tape measure proteins); and (iii) *Burkholderia* phage KS9 (HNH, TerL, portal, protease/scaffold, major capsid, head closure). Among these, the KS9 and Dina phages were also the ones with which VarioGold created a clade in the ViPTree analysis based on the limited similarity between regions encoding the mentioned proteins (Figure S2). These three phages have not been completely taxonomically classified. Marilyn (MT448615) and KS9 (FJ982340) are without not only genus-level but also family-level taxonomy information, while phage Dina (MT740734) was classified as the only representative of the Dinavirus genus [56]; moreover, the VarioGold genome does not share any nucleotide sequence similarity with the Dina, Marilyn and KS9 genomes based on both BLASTN searches and OrthoANLu [57] analyses. In the case of the latter, the only similarity was observed between VarioGold and Dina phages across 669 bp with 58.5% OrthoANLu values. These results suggest the lack of a significant phylogenetic relationship between VarioGold and known phages, thus indicating that VarioGold is a novel virus; thus, regarding current taxonomic assignment procedures, it may be considered a new viral genus [54,58,59].



**Figure 3.** Comparative genome alignment of VarioGold and the three phages. The alignment was created with Clinker using default settings. Each ORF is represented by an arrow. ORF-encoding proteins that did not share sequence similarity are colored gray, while others are connected with blocks reflecting the degree of sequence identity and are color-coded. Sequence identity ranged between 32% and 60%.



### 2.5.2. Identification and General Genomic Features of *Variovorax* spp. Prophages

Since VarioGold is the first bacteriophage of *Variovorax* spp., no comparisons could be made to other *Variovorax* phages; therefore, we attempted to identify prophage sequences in complete *Variovorax* spp. genomes that were extracted from the NCBI database.

As of 25 October 2021, 21 publicly available complete genomes of the *Variovorax* genus (Table S2) were available and these were searched for the presence of putative prophages, as previously reported [60,61]. Briefly, for the initial integrase and attL/attR prediction, the PhiSpy algorithm was used [39] and manually inspected by assessing the phage localization in the host genome. Subsequently, we verified the presence of the essential structure and packaging genes using Virfam (<http://biodev.cea.fr/virfam/>, accessed on 17 January 2022) and BLASTP [62]. Only the regions containing complete prophage genomes were included in further analyses. All prophage elements lacking matches to core phage proteins (e.g., terminase, capsid, head, tail proteins) were excluded. Genome annotation was further verified using the MAISEN web service in order to determine the genomic context of the investigated genes [63].

As a result, 37 novel prophages were identified (Table 2). These were detected in almost all *Variovorax* spp., except *Variovorax* sp. PAMC-28562, PBL-H6 and SRS16. Among them, 10 out of 21 were found to be polylysogenic, carrying multiple prophages within their genomes: five strains had more than two complete prophage regions, of which two strains (*Variovorax* sp. PMC12 and *V. paradoxus* VAI-C) had four such sequences, and a single strain (*V. boronicumulans* J1) had five. Virfam analysis indicated that, presumably, there were 17 of siphoviral morphology, 16 of podoviral morphology and 4 of myoviral morphology. The integration modules of the identified prophages encoded tyrosine recombinases (33 prophages), serine recombinases (3) or Mu-like transposases (2) (Table 2).

Putative integration sites were identified for the majority of the tyrosine recombinase-encoding *Variovorax* viruses (Table 2). For 28 prophages, these sites were various tRNA genes, of which the most commonly targeted were (i) tRNA<sub>Arg</sub>, which was used by seven prophages, and (ii) tRNA<sub>Ser</sub>, each of which were used by nine prophages and also by VarioGold. These observations corroborate previous findings regarding the preferential integration of phages (and other integrative elements) within tRNA genes [64].

The genome size of the identified prophages ranged from 38 kb to 113 kb. Total prophage genomes accounted for 0.5–6% of the bacterial chromosome, which appeared to be lower compared to that of other bacterial genomes (10–20%). As examples, the genomes of the *Streptococcus pyogenes* and *Escherichia coli* O157:H7 strains contain 12% and 16% prophage sequences, respectively [65]. The genomes of identified prophages were aligned with Clinker to explore the similarities in their structures and encoded proteins (Figure S3). The analysis revealed high mosaicism of their genomes and sets of unique proteins encoded by each prophage. The most distinctive ones were 5C-2\_pp\_1, HW608\_pp\_2, PBS-H4\_pp\_1 and PBS-H4\_pp\_2, which shared—at most—seven proteins (based on at least 30% sequence identity) with other prophages. Despite that, eight groups of prophages encoding more similar proteins could be indicated.

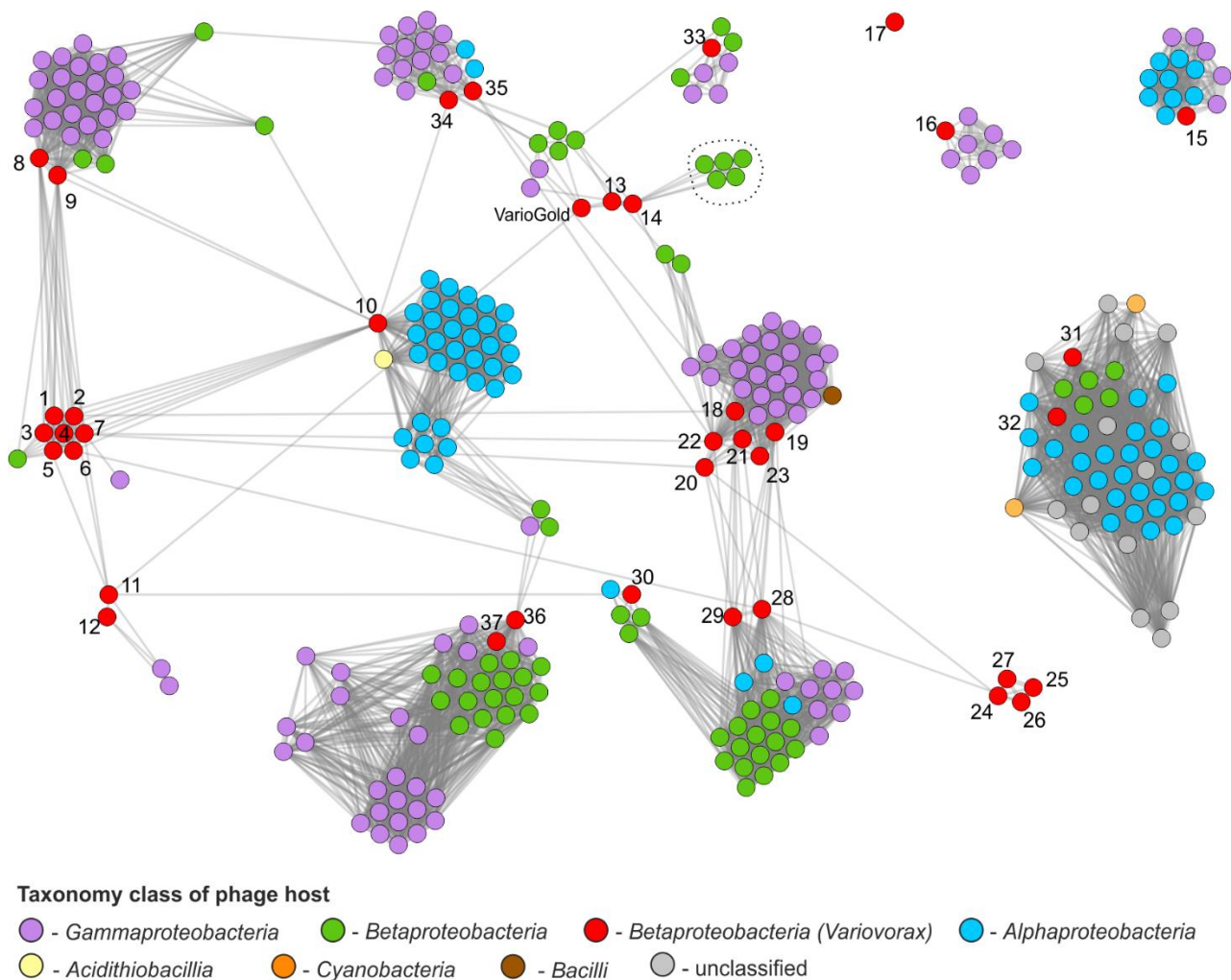
**Table 2.** Characteristics of Variovorax prophages identified in the genomic sequences from the NCBI database.

| No. | Prophage        | Strain (Accession No)                       | Coordinates          | Virfam                 | Site of Integration                          | att Sequence  | Type of Integrase   | Genome Size (bp) |
|-----|-----------------|---|----------------------|------------------------|--|---|---------------------|------------------|
| 1.  | PMC12_pp_2      | <i>Variovorax</i> sp. PMC12 (CP027773)      | 5,166,871..5,204,531 | Siphoviridae of Type 1 | tRNA-Ser(TGA)                                | TCTCACACTCTCCGCCAGAATCAATC(T/C)(T/C)<br>TGGCAGTTTTTGAAA-GTCCCGCGCAGCCTCTGCGA        | Tyrosine            | 37,661           |
| 2.  | VAI-C_pp_1      | <i>V. paradoxus</i> VAI-C (CP063166)        | 5,357,355..5,405,518 | Siphoviridae of Type 1 | tRNA-Ser(CGA)                                | TCCCTCCCTCTCTCTCAA  | Tyrosine            | 48,164           |
| 3.  | B4_pp_1         | <i>V. paradoxus</i> B4 (CP003911)           | 4,696,119..4,740,974 | Siphoviridae of Type 1 | tRNA-Ser(TGA)                                | CACACTCTCCGCCAGAATCCAATCTTTGGCAGTTTCTTG<br>AAGTCCCGCGCAGTTCATGCGACGGGACTTTTTCATTGGG | Tyrosine            | 44,856           |
| 4.  | CSUSB_pp_1      | <i>V. paradoxus</i> CSUSB (CP046622)        | 3,365,369..3,408,809 | Siphoviridae of Type 1 | tRNA-Ser(GCT)                                | CCTCCGTTCCGCCAA   | Tyrosine            | 43,441           |
| 5.  | J1_pp_1         | <i>V. boronicumulans</i> J1 (CP023284)      | 2,675,825..2,721,941 | Siphoviridae of Type 1 | tRNA-Val(TAC)                                | CCCTTACAAGGCGTAGGTCTGGGGGTTTC<br>GAGCCCTCAGCACCCACCACCA                             | Tyrosine            | 46,117           |
| 6.  | J1_pp_4         | <i>V. boronicumulans</i> J1 (CP023284)      | 4,415,582..4,454,484 | Siphoviridae of Type 1 | tRNA dihydrouridine synthase DUS4            | GCGCTCGCTCGGG   | Tyrosine            | 38,903           |
| 7.  | vvax_pp_1       | <i>V. paradoxus</i> vvax (LR743507)         | 4,988,968..5,029,379 | Siphoviridae of Type 1 | tRNA-Ser(TGA)                                | CTCGCGCAACCA  | Tyrosine            | 40,412           |
| 8.  | PAMC 26660_pp_1 | <i>Variovorax</i> sp. PAMC 26660 (CP060295) | 1,178,078..1,222,407 | Siphoviridae of Type 1 | Sigma-70 family RNA polymerase               | GTTGCCAGCTTCTTGCGCAGCCACGACT<br>GGAGCCAGCCGTGGTGGTCGC                               | Transposase Mu-like | 44,330           |
| 9.  | RKNM96_pp_1     | <i>Variovorax</i> sp. RKNM96 (CP046508)     | 6,243,874..6,284,473 | Siphoviridae of Type 1 | Intergenic region                            | TG ... CA   | Transposase Mu-like | 40,600           |
| 10. | HW608_pp_3      | <i>Variovorax</i> sp. HW608 (LT607803)      | 7,244,519..7,304,853 | Siphoviridae of Type 1 | Intergenic region                            | Not identified  | Serine              | 60,335           |
| 11. | PAMC 28711_pp_2 | <i>Variovorax</i> sp. PAMC 28711 (CP014517) | 234,890..279,029     | Siphoviridae of Type 1 | Flavin reductase                             | ATGGACATCGACTTCGCCACCCTCACCGAATACCAGCGCTACAA  | Tyrosine            | 44,140           |
| 12. | VAI-C_pp_2      | <i>V. paradoxus</i> VAI-C (CP063166)        | 4,386,549..4,428,859 | Siphoviridae of Type 1 | DNA competence protein ComEC/Rec2            | GCTGCCGTGGTGGCGC  | Tyrosine            | 42,311           |
| 13. | J1_pp_2         | <i>V. boronicumulans</i> J1 (CP023284)      | 349,148..353,129     | Siphoviridae of Type 1 | 30S ribosomal S12 methylthiotransferase RimO | GTCGCCGTCTTGCGC   | Serine              | 39,810           |
| 14. | J1_pp_3         | <i>V. boronicumulans</i> J1 (CP023284)      | 4,205,701..4,263,544 | Siphoviridae of Type 1 | tRNA-Arg(TCT)                                | ATCCCTCCCG  | Tyrosine            | 52,010           |
|     | VarioGold       | <i>Variovorax</i> sp. ZS18.2.2              | -                    | Siphoviridae of Type 1 | tRNA-Ser(CGA)                                | CCTCCCTCTCTCCA  | Tyrosine            | 39,429           |
| 15. | 5C-2_pp_1       | <i>V. paradoxus</i> 5C-2 (CP045644)         | 3,926,575..4,039,667 | Siphoviridae of Type 1 | tRNA-Pro(GGG)                                | TTGCATGGGGTGCAAGGGGTGCAAGGTTCGAATCCTTT<br>CACACCG-ACCAATAA                          | Tyrosine            | 113,093          |
| 16. | PBS-H4_pp_2     | <i>Variovorax</i> sp. PBS-H4 (LR594675)     | 3,082,071..3,170,589 | Siphoviridae of Type 1 | tRNA-Gly(CCC)                                | GTCTACCATTTGAACACCCGCA  | Tyrosine            | 88,469           |
| 17. | HW608_pp_2      | <i>Variovorax</i> sp. HW608 (LT607803)      | 5,030,346..5,096,253 | Siphoviridae of Type 1 | peptidylprolyl isomerase                     | TCCATACGAGAATTC-TCC   | Tyrosine            | 60,846           |
| 18. | PMC12_pp_3      | <i>Variovorax</i> sp. PMC12 (CP027773)      | 5,692,545..5,758,265 | Podoviridae of Type 3  | Intergenic region                            | CTGGCTACCCG(C/G)CT(A/G)GCTACCC  | Tyrosine            | 65,721           |

Table 2. Cont.

| No. | Prophage        | Strain (Accession No)                       | Coordinates              | Virfam                | Site of Integration | att Sequence   | Type of Integrase | Genome Size (bp) |
|-----|-----------------|---|--------------------------|-----------------------|---------------------|--|-------------------|------------------|
| 19. | PDNC026_pp_1    | <i>Variovorax</i> sp. PDNC026 (CP070343)    | 5,173,628..<br>5,241,433 | Podoviridae of Type 3 | tRNA-His(GTG)       | CAGATTGTGATTCTGGTCGTCGGGTTTC<br>GAGTCCCATCAGCCACCCCAA          | Tyrosine          | 67,886           |
| 20. | PMC12_pp_4      | <i>Variovorax</i> sp. PMC12 (CP027773)      | 2,605,254..<br>2,666,437 | Podoviridae of Type 3 | tRNA-Leu(CAA)       | TGTGGTGCCCGGGCCGAATCGAACCGG<br>CACACCTTTCGGTGGGGGATTTTGAAGTCCC | Tyrosine          | 61,184           |
| 21. | EPS_pp_1        | <i>V. paradoxus</i> EPS (CP002417)          | 2,169,141..<br>2,234,652 | Podoviridae of Type 3 | tRNA-His(GTG)       | CAGATTGTGATTCTGGTCGTCGGGTTTC<br>GAGTCCCATCAGCCACCCCAA          | Tyrosine          | 65,512           |
| 22. | VAI-C_pp_3      | <i>V. paradoxus</i> VAI-C (CP063166)        | 2,942,097..<br>3,007,349 | Podoviridae of Type 3 | tRNA-Asn(GTT)       | TGGCTCCTCGACCTGGGCTCGAACCAGGGA-<br>CCTACGGATTAAACAGTC          | Tyrosine          | 65,253           |
| 23. | 38R_pp_1        | <i>Variovorax</i> sp. 38R (CP062121)        | 3,617,751..<br>3,688,219 | Podoviridae of Type 3 | tRNA-Arg(TCT)       | Not identified   | Tyrosine          | 70,469           |
| 24. | PAMC 28711_pp_1 | <i>Variovorax</i> sp. PAMC 28711 (CP014517) | 42,947..<br>81,571       | Podoviridae of Type 3 | tRNA-Arg(TCT)       | TGGCCTGTCCGGAGGGGATCGAACCCCGAC<br>AACCTGCTTAGAAGGCAG           | Tyrosine          | 38,625           |
| 25. | RA8_pp_1        | <i>Variovorax</i> sp. RA8 (LR594662)        | 5,162,402..<br>5,202,563 | Podoviridae of Type 3 | tRNA-Arg(ACG)       | GGCTACGAACCAAGGGGTCGTGGGTTTCAAT<br>CCTGCCAGCCGCACCACTTTT       | Tyrosine          | 40,162           |
| 26. | PBL-E5_pp_1     | <i>Variovorax</i> sp. PBL-E5 (LR594671)     | 4,567,056..<br>4,606,930 | Podoviridae of Type 3 | tRNA-Arg(CCT)       | TGGTGCCCTCGACAGGAATCGAACCTG                                    | Tyrosine          | 39,875           |
| 27. | WDL1_pp_1       | <i>Variovorax</i> sp. WDL1 (LR594689)       | 2,078,592..<br>2,121,776 | Podoviridae of Type 3 | tRNA-Arg(CCT)       | AGGTTTCGATTCTGTCTGAGGGCACCAGTAAGGT                             | Tyrosine          | 43,185           |
| 28. | PBS-H4_pp_3     | <i>Variovorax</i> sp. PBS-H4 (LR594675)     | 4,606,795..<br>4,670,479 | Podoviridae of Type 3 | tRNA-His(GTG)       | CAGATTGTGATTCTGGTCGTCGGGTTTC<br>GAGTCCCATCAGCCACCCCAA          | Tyrosine          | 63,685           |
| 29. | PAMC 26660_pp_2 | <i>Variovorax</i> sp. PAMC 26660 (CP060295) | 4,885,415..<br>4,937,289 | Podoviridae of Type 3 | tRNA-His(GTG)       | CAGATTGTGATTCTGGTCGTCGTGGGTTTCGAGT<br>CCCATCAGCCACCCCAA        | Tyrosine          | 51,952           |
| 30. | VAI-C_pp_4      | <i>V. paradoxus</i> VAI-C (CP063166)        | 3,837,980..<br>3,880,010 | Podoviridae of Type 3 | tRNA-Ser(GCT)       | TTGGCGGAACCGGAGG   | Tyrosine          | 42,031           |
| 31. | PMC12_pp_1      | <i>Variovorax</i> sp. PMC12 (CP027773)      | 1,531,523..<br>1,574,372 | Podoviridae of Type 3 | tRNA-Leu(TAA)       | TTCGGGGCACCA   | Tyrosine          | 42,264           |
| 32. | RKNM96_pp_2     | <i>Variovorax</i> sp. RKNM96 (CP046508)     | 4,013,608..<br>4,057,349 | Podoviridae of Type 3 | tRNA-Asn(GTT)       | ACTGTTAATCCGTAGGTCCTGGTTCGAGCCCAGGTCGAGGAGCCA                  | Tyrosine          | 43,742           |
| 33. | PBS-H4_pp_1     | <i>Variovorax</i> sp. PBS-H4 (LR594675)     | 1,630,279..<br>1,670,591 | Podoviridae of Type 3 | tRNA-Ser(CGA)       | TCCCACCCTCTCCGCCAGCA   | Tyrosine          | 40,313           |
| 34. | PAMC 26660_pp_3 | <i>Variovorax</i> sp. PAMC 26660 (CP060295) | 4,563,808..<br>4,646,346 | Myoviridae of Type 1  | Intergenic region   | CGGGGGTTCAAATCCCCCA  | Tyrosine          | 82,539           |
| 35. | WDL1_pp_2       | <i>Variovorax</i> sp. WDL1 (LR594689)       | 662,952..<br>719,595     | Myoviridae of Type 1  | tRNA-Ser(ACT)       | GTAGTGGCTCCTCGACCTGGGCTCGAA<br>CCAGGGACCTACGGATTAAACAG         | Tyrosine          | 56,644           |
| 36. | J1_pp_5         | <i>V. boronicumulans</i> J1 (CP023284)      | 6,754,149..<br>6,793,076 | Myoviridae of Type 1  | tRNA-Met(CAT)       | TGGTTGCGCGAG   | Tyrosine          | 38,928           |
| 37. | PDNC026_pp_2    | <i>Variovorax</i> sp. PDNC026 (CP070343)    | 4,474,061..<br>4,514,959 | Myoviridae of Type 1  | tRNA-Arg(TCT)       | TTGGCCTGCCCGAGGGGATCGAACC                                      | Serine            | 40,899           |

To better understand the diversity and relationships among prophages of *Variovorax* spp., protein-sharing networks were generated with vConTACT2 and the INPHARED database (1 December 2021 release, over 17,470 phage genomes). In the resulting network, *Variovorax* prophages demonstrated significant proteome similarity to, in total, 280 known phages (Figures 4 and S4). Prophages were spread across various clusters, except for one—HW608\_pp\_2—which remained a singleton. They were often located outside clusters and acted as bridges between them, thereby reflecting their mosaic genome structure. Interestingly, the prophages present in the same strain (J1\_pp1 and J1\_pp4; J1\_pp2 and J1\_pp3; PMC12\_pp3 and PMC12\_pp4) are similar to each other and clustered together, which is not common.



**Figure 4.** Proteome-based similarity network. The network was constructed with vConTACT2 to explore the diversity of *Variovorax* prophages in comparison with other known phages. Each node represents a single phage genome, and the edge represents a significant similarity between proteomes of connected phages. Dashed lines surrounding five phages indicate *Inoviridae* phages RSS0, RSS1, RSS-TH1, RS611 and RSBg infecting *Ralstonia* spp. Numbers next to *Variovorax* prophages correspond to numbering in Table 2.

According to vConTACT2, 7 *Variovorax* prophages were considered to be outliers (PBS-H4\_pp\_1, PBS-H4\_pp\_2, HW608\_pp\_2, HW608\_pp\_3, 5C-2\_pp\_1, VAI-C\_pp4, J1\_pp5) and 3 were considered clustered/singletons (PAMC28711\_pp2, PMC12\_pp1, VAI-C\_pp2), while 18 form 5 *Variovorax*-specific viral clusters (including the one with VarioGold; see below), 5 form extended viral clusters with known viruses (PMC12\_pp\_3 and PDNC026\_pp\_1;

WDL1\_pp\_2 and PAMC26660\_pp\_3; PDNC026\_pp\_2), and others overlapped multiple viral clusters.

The highest similarities were found for seven siphoviral prophages, i.e., PMC12\_pp\_2, VAI-C\_pp\_1, B4\_pp\_1, CSUSB\_pp\_1, vvax\_pp\_1, J1\_pp\_1 and J1\_pp\_4, which cluster within the same clique (numbers 1–7 in Figure 4). They also shared similarities in several tail proteins with almost identical Mu-like prophages PAMC 26660\_pp\_1 and RKNM96\_pp\_1, which were also predicted by Virfam as siphoviruses (numbers 8–9 in Figure 4). All of them were connected with siphovirus HW608\_pp3 (number 10 in Figure 4).

Another group is formed by four prophages with predicted podoviral morphology, PAMC 28711\_pp\_1, RA8\_pp\_1, PBL-E5\_pp\_1 and WDL1\_pp\_1 (numbers 24–27 in Figure 4), which shared, among other characteristics, a similar integration site in the tRNA<sub>Arg</sub> gene (Table 2).

It is also worth noting that prophages PMC12\_pp\_1 and RKNM96\_pp\_2 (numbers 31–32 in Figure 4) are a part of a dense clique that gathers representatives of the *Autographiviridae* family and are characterized by encoding their own single subunit RNA polymerase (RNAP) and a common unidirectional gene arrangement. Both these features are shared with this family by the PMC12\_pp\_1 and RKNM96\_pp\_2 prophages. The RNAP (AVQ80729.1) of PMC12\_pp\_1 showed a 56% identity with its counterparts of virulent *Ralstonia* phages (e.g., P-PSG-11, [66]) and *Bordetella* phage vB\_BbrP\_BB8 (QDB70995.1). The putative RNAP of RKNM96\_pp\_2 shares a 36% identity with the RNAP encoded by *Rhizobium* phages vB\_RleA\_TRX32-1 and RHEph01 [67] and also the RNAP of the temperate *Teseptimavirus* S2B-infecting *Caulobacter crescentus* CB15 [68]. Until recently, T7-like phages were considered strictly virulent. Nevertheless, several putative representatives of the *Autographiviridae* family have been identified lately in several Gram-negative and several Gram-positive bacterial genomes, or ones that are able to establish a lysogenic relationship with their host—for example, *Teseptimavirus* S2B, which is mentioned above. Therefore, prophages PMC12\_pp\_1 and RKNM96\_pp\_2 would be other examples of these.

A somewhat surprising result of the vConTACT2 network analysis is the presence of a group of single-stranded *Inoviridae* phages that share three proteins with the J1\_pp3 prophage, which were predicted to be a zonular occludens toxin, minor coat protein and attachment protein (ATA54428.1, ATA54427.1, ATA54426.1, respectively); the consecutive genes that encode them form a segment that interrupts the structural module of the J1\_pp3 prophage.

VarioGold showed the highest similarity with prophages J1\_pp2, VAI-C\_pp2 and J1\_pp3, sharing 19, 8 and 7 encoded proteins with them, respectively (Figure S5); moreover, 17 protein products of VarioGold genes could be considered unique as they did not show any similarities with proteins encoded by the *Variovorax* prophages identified by us. These are proteins encoded by putative early genes (VG\_p30–34 and VG\_p40–41), a lysogenic switch module (CI, Cro, CII), TerS, structural proteins (VG\_p06–07), the host recognition–lysis module (tail–spike, lysozyme) and the moron-like segment of the right end of the VarioGold prophage sequence (VG\_p25–p28; see above).

### 3. Materials and Methods

#### 3.1. Bacterial Strains, Media and Growth Conditions

For bacterial isolation, rock biofilm samples of the Zloty Stok gold and arsenic mine (SW Poland), which were collected in February 2021, were mixed vigorously in 0.7% NaCl and then serially diluted in saline. Aliquots of 100 µL of each dilution were plated on Reasoner's 2A (R2A) medium, solidified with 1.5% (*w/v*) agar [69] and incubated at 17 °C for 7 days, followed by incubation at 6 °C for a further two weeks. The obtained colonies were used to isolate pure cultures. *Variovorax paradoxus* EPS [13] and *Variovorax paradoxus* B4 [70], used in host range testing, were routinely grown under aerobic conditions in R2A medium at 20 °C.



For the temperature tolerance analysis, *Variovorax* sp. ZS18.2.2 was plated with the use of the streak plate technique. The plates were incubated for 7 days at various temperatures ranging from 4–37 °C and examined after every 24 h. The data were obtained from three independent experiments.

### 3.2. Determination of the Minimum Inhibitory Concentrations of Arsenite and Arsenate

The minimum inhibitory concentrations (MIC) of arsenite and arsenate were established using the broth dilution method. Sterile tubes (15 mL) containing R2A medium amended with the respective amount of arsenite/arsenate were inoculated with overnight cultures to a final optical density at 600 nm ( $OD_{600}$ ) of 0.045 and incubated for 48 h at 20 °C with shaking (150 RPM). The optical density was measured immediately after inoculation and every 24 h with an automated plate reader (Sunrise, TECAN, Männedorf, Switzerland). The following arsenic compounds were used for MIC determination:  $NaAsO_2$  (0–50 mM) and  $Na_2HAsO_4$  (0–500 mM). The MIC was defined as the lowest concentration of  $As^{n+}$  that completely inhibited the growth of bacteria. The data were obtained from three independent experiments.

### 3.3. Oxidation and Reduction of Arsenic Compounds and Arsenic Speciation Assay

In order to investigate the strain's potential for arsenic oxidation and reduction, aerobic cultures in the R2A medium amended with the respective amount of arsenite/arsenate were set as previously performed in the MIC experiment (see Section 3.2) and incubated for 48 h at 20 °C with shaking (150 RPM). Arsenic species present in culture supernatant after 48 h of incubation were determined by the silver nitrate test described by Drewniak et al. [23]. Supernatant from each culture (200  $\mu$ L) was taken and mixed at a ratio of 1:20 with a 50 mM silver nitrate (V) solution. The addition of  $AgNO_3$  solution to the culture containing As(III) compounds caused the precipitation of a yellow  $Ag_3AsO_3$  (silver arsenite). The presence of As(V) compounds caused the precipitation of a brown  $Ag_3AsO_4$  (silver arsenate).

### 3.4. Standard Molecular Biology Procedures

Standard DNA manipulations were carried out according to the protocols described by Sambrook and Russell [69].

### 3.5. Induction and Isolation of Phage Particles

To induce a potential prophage in *Variovorax* sp. ZS18.2.2, bacterial cells were treated with mitomycin C (500 ng/mL, MilliporeSigma, Darmstadt, Germany), and their growth (with shaking) was continued for 18 h. The resulting lysate was cleared of cell debris by centrifugation (13,000 RPM, 30 min). The supernatant was condensed on an Amicon ultra-filtration column Ultracel-100K (Merck Millipore, Ireland) and used for further analysis.

### 3.6. DNA Isolation and Sequencing and Bioinformatics

The total DNA was isolated from *Variovorax* sp. ZS18.2.2 using a Genomic Mini Kit (A&A Biotechnology, Gdansk, Poland). The whole-genome shotgun sequencing of the ZS18.2.2 strain was conducted by Eurofins Genomics Germany GmbH (Ebersberg, Germany) on an Illumina NovaSeq6000 platform at a read length of  $2 \times 150$  bp.

DNA of the VarioGold phage was isolated by phenol–chloroform extraction and isopropanol precipitation [69], and its sequencing was also performed by Eurofins Genomics with the same parameters as for bacterial genomic DNA.

The raw reads acquired from the above sequencing projects were subjected to a quality check and filtering with the application of FastQC v.0.11.5 [71] and fastp v.0.21.0 [72]. During the fastp run, the following parameters were applied: `–detect_adapter_for_pe` `–cut_window_size 8` `–cut_tail` `–cut_mean_quality 24` `–length_required 50` `–length_limit 160` `–n_base_limit 5` `–trim_poly_x` `–poly_x_min_len 0` `–correction` `–overlap_len_require 20` `–overlap_diff_limit 5`. The filtered reads were then assembled with SPAdes v.3.15.3 in

an isolated mode with the following kmers: 33, 55, 77, 99 and 127 [73]. The analysis of the genomes' sequence coverage, including the analysis of the redundant regions, was performed by mapping the filtered reads against the assemblies with bwa mem v.0.7.17-r1198-dirty [74] and samtools v.1.10 [75]. Then, the alignments were viewed in Integrative Genome Viewer v.2.6.2 [76]. Additional analysis of the phage termini and packaging mechanisms was conducted with the application of PhageTerm v.1.0.12, using the filtered reads as the input [40].

### 3.7. Transmission Electron Microscopy (TEM)

TEM analysis was conducted as described previously [77]. The visualization of the phages was performed at the Core Facility of the International Institute of Molecular and Cell Biology (IIMCB, Warsaw, Poland).

### 3.8. Genome Annotation

The analysis of the nucleotide sequence of the VarioGold phage was performed using Clone Manager 8 (Sci-Ed) and Artemis v18.0.0 software [78]. The genomes were automatically annotated using the RASTtk [79] in phage mode on the PATRIC website [80]. The annotations were manually verified based on homology searches using BLAST programs against the NCBI non-redundant (nr) and SwissProt databases [62], HHpred using HHpred or HMMER tools against the PDB\_mmCIF70\_11\_Oct, SCOPe70\_2.07, COG\_KOG\_v.1.0 and Pfam-A\_v35; and NCBI\_Conserved\_Domains (CD)\_v3.18 [81], InterProScan v5.48 [82], Pfam [83] and UniProt [84]. The transmembrane helices were identified with the help of the TMHMM v2.0 server [85]. Putative tRNA genes were identified using the tRNAScan-SE [86] and ARAGORN programs [87]. A phage morphotype search was carried out using Virfam [88]. Phage termini were predicted using PhageTerm [40]. All software was run at default settings.

### 3.9. Comparative Genomics Analysis

Comparative genomics analysis of the phage genomes was performed with the application of Clinker using the default settings [89]. If necessary, the genomes were circulated and re-oriented to enhance the overview of genome structure conservation. The genome of the VarioGold phage was also compared with other known phage genomes recovered from the INPHARED database (as of the 1 December 2021 release) [90] with the application of vConTACT2 v0.9.20, as described previously [54]. All of the analyzed networks were visualized with Gephi v.0.9.2 [91], and the nodes were laid out in two-dimensional space with the application of the ForceAtlas 2 [92] and Noverlap algorithms.

For phylogenetic analysis, the VarioGold phage sequence was uploaded to the ViPTree server updated on 20 June 2022 [55]. The analysis was run against all dsDNA prokaryotic viruses with automatic gene prediction. Then, the resulting tree was recalculated with a subset of neighbor phages. Selected genomes were compared with VarioGold using BLASTN (with e-value lower than 0.1 threshold) and OrthoANIu [57] to determine its taxonomic membership.

### 3.10. Nucleotide Sequence Accession Numbers

The draft genome of *Variovorax* sp. ZS18.2.2, as well as the VarioGold phage genome, can be accessed under the following GenBank accession Nos: JANTLNM000000000 and OP296522, respectively.

## 4. Conclusions

In this study, we identified and characterized the first *Variovorax* virus—an inducible temperate phage, VarioGold—residing in the *Variovorax* sp. strain ZS18.2.2's genome, which was isolated from a biofilm collected from the Zloty Stok gold and arsenic mine (Poland). The slight resemblance of VarioGold to other known phages at both the nucleotide and protein levels suggests that it should be considered a new viral genus of the *Caudoviricetes* class.

We performed an insightful analysis of 21 publicly available *Variovorax* complete genomes, and in 18 of them, an additional 37 complete prophage sequences were identified. It is striking that almost all of the analyzed *Variovorax* genomes carry at least one prophage, and as many as ten of them carry more than one, which allows us to conclude that polylysogeny is common in the *Variovorax* genus; moreover, a protein-based similarity network showed the high diversity of these phages. Finally, a global analysis of the identified (pro)phages with known viruses revealed that they show a diversified level of similarity to them, often acting as bridges in network analyses, which revealed that they are partly similar to the linked groups of highly similar phages, thereby filling the gaps of viral dark matter. This work significantly expands current knowledge on the diversity of bacteriophages that infect *Betaproteobacteria*; moreover, it provides great evidence regarding how much information about the viral ‘dark matter’ can be obtained by conducting simple and low-cost analyses of bacterial genomic sequences that are deposited in public databases.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms232113539/s1>. References [93–96] are cited in the supplementary materials.

**Author Contributions:** Conceptualization, M.R. and P.D.; methodology, M.R. and P.D.; software, P.D.; validation, M.R. and P.D.; formal analysis, M.R.; investigation, M.R., P.D. and M.K.; resources, M.R.; data curation, M.R., P.D. and M.K.; writing—original draft preparation, M.R., P.D. and M.K.; visualization, P.D. and M.R.; supervision, M.R.; project administration, M.R.; funding acquisition, M.R. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Science Centre (Poland) within the project grant no. 2017/25/B/NZ8/00472 (to M.R.).

**Acknowledgments:** We would like to thank Matylda Macias (the Core Facility of IIMCB) for support with TEM.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Satola, B.; Wübbeler, J.H.; Steinbüchel, A. Metabolic characteristics of the species *Variovorax paradoxus*. *Appl. Microbiol. Biotechnol.* **2013**, *97*, 541–560. [CrossRef] [PubMed]
2. Anesti, V.; McDonald, I.R.; Ramaswamy, M.; Wade, W.G.; Kelly, D.P.; Wood, A.P. Isolation and molecular detection of methylo-trophic bacteria occurring in the human mouth. *Environ. Microbiol.* **2005**, *7*, 1227–1238. [CrossRef]
3. Ciok, A.; Dziewit, L.; Grzesiak, J.; Budzik, K.; Gorniak, D.; Zdanowski, M.K.; Bartosik, D. Identification of miniature plasmids in psychrophilic Arctic bacteria of the genus *Variovorax*. *FEMS Microbiol. Ecol.* **2016**, *92*, fiw043. [CrossRef] [PubMed]
4. Breugelmans, P.; D’Huys, P.J.; De Mot, R.; Springael, D. Characterization of novel linuron-mineralizing bacterial consortia enriched from long-term linuron-treated agricultural soils. *FEMS Microbiol. Ecol.* **2007**, *62*, 374–385. [CrossRef] [PubMed]
5. Benedek, T.; Szentgyörgyi, F.; Gergőcs, V.; Menashe, O.; Gonzalez, P.A.F.; Probst, A.J.; Kriszt, B.; Táncsics, A. Potential of *Variovorax paradoxus* isolate BFB1\_13 for bioremediation of BTEX contaminated sites. *AMB Express* **2021**, *11*, 126. [CrossRef]
6. Futamata, H.; Nagano, Y.; Watanabe, K.; Hiraishi, A. Unique kinetic properties of phenol-degrading *variovorax* strains responsible for efficient trichloroethylene degradation in a chemostat enrichment culture. *Appl. Environ. Microbiol.* **2005**, *71*, 904–911. [CrossRef]
7. Liu, Z.H.; Cao, Y.M.; Zhou, Q.W.; Guo, K.; Ge, F.; Hou, J.Y.; Hu, S.Y.; Yuan, S.; Dai, Y.J. Acrylamide biodegradation ability and plant growth-promoting properties of *Variovorax boronicumulans* CGMCC 4969. *Biodegradation* **2013**, *24*, 855–864. [CrossRef]
8. Murdoch, R.W.; Hay, A.G. The biotransformation of ibuprofen to trihydroxyibuprofen in activated sludge and by *Variovorax* Ibu-1. *Biodegradation* **2015**, *26*, 105–113. [CrossRef]
9. Leadbetter, J.R.; Greenberg, E.P. Metabolism of acyl-homoserine lactone quorum-sensing signals by *Variovorax paradoxus*. *J. Bacteriol.* **2000**, *182*, 6921–6926. [CrossRef]
10. Sun, S.L.; Yang, W.L.; Fang, W.W.; Zhao, Y.X.; Guo, L.; Dai, Y.J. The Plant Growth-Promoting Rhizobacterium *Variovorax boronicumulans* CGMCC 4969 Regulates the Level of Indole-3-Acetic Acid Synthesized from Indole-3-Acetonitrile. *Appl. Environ. Microbiol.* **2018**, *84*, e00298–18. [CrossRef]
11. Vurukonda, S.S.; Vardharajula, S.; Shrivastava, M.; SkZ, A. Enhancement of drought stress tolerance in crops by plant growth promoting rhizobacteria. *Microbiol. Res.* **2016**, *184*, 13–24. [CrossRef] [PubMed]
12. Satsuma, K. Mineralisation of the herbicide linuron by *Variovorax* sp. strain RA8 isolated from Japanese river sediment using an ecosystem model (microcosm). *Pest Manag. Sci.* **2010**, *66*, 847–852. [CrossRef]

13. Han, J.I.; Spain, J.C.; Leadbetter, J.R.; Ovchinnikova, G.; Goodwin, L.A.; Han, C.S.; Woyke, T.; Davenport, K.W.; Orwin, P.M. Genome of the Root-Associated Plant Growth-Promoting Bacterium *Variovorax paradoxus* Strain EPS. *Genome Announc.* **2013**, *1*, e00843-13. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Han, S.R.; Lee, J.H.; Kang, S.; Park, H.; Oh, T.J. Complete genome sequence of opine-utilizing *Variovorax* sp. strain PAMC28711 isolated from an Antarctic lichen. *J. Biotechnol.* **2016**, *225*, 46–47. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Belimov, A.A.; Dodd, I.C.; Hontzeas, N.; Theobald, J.C.; Safronova, V.I.; Davies, W.J. Rhizosphere bacteria containing 1-aminocyclopropane-1-carboxylate deaminase increase yield of plants grown in drying soil via both local and systemic hormone signalling. *New Phytol.* **2009**, *181*, 413–423. [\[CrossRef\]](#)
16. Belimov, A.A.; Hontzeas, N.; Safronova, V.I.; Demchinskaya, S.V.; Piluzza, G.; Bullitta, S.; Glick, B.R. Cadmium-tolerant plant growth-promoting bacteria associated with the roots of Indian mustard. *Soil Biol. Biochem.* **2005**, *37*, 41–250. [\[CrossRef\]](#)
17. Tamburini, E.; Sergi, S.; Serreli, L.; Bacchetta, G.; Milia, S.; Cappai, G.; Carucci, A. Bioaugmentation-Assisted Phytostabilisation of Abandoned Mine Sites in South West Sardinia. *Bull. Environ. Contam. Toxicol.* **2016**, *98*, 310–316. [\[CrossRef\]](#)
18. Terry, L.R.; Kulp, T.R.; Wiatrowski, H.; Miller, L.G.; Oremland, R.S. Microbiological Oxidation of Antimony(III) with Oxygen or Nitrate by Bacteria Isolated from Contaminated Mine Sediments. Applied and Environmental Microbiology. *Appl. Environ. Microbiol.* **2015**, *81*, 8478–8488. [\[CrossRef\]](#)
19. Piotrowska-Seget, Z.; Cycoń, M.; Kozdrój, J. Metal-tolerant bacteria occurring in heavily polluted soil and mine spoil. *Appl. Soil Ecol.* **2005**, *28*, 237–246. [\[CrossRef\]](#)
20. Szentgyörgyi, F.; Benedek, T.; Fekete, D.; Táncsics, A.; Harkai, P.; Kriszt, B. Development of a bacterial consortium from *Variovorax paradoxus* and *Pseudomonas veronii* isolates applicable in the removal of BTEX. *AMB Express* **2022**, *12*, 4. [\[CrossRef\]](#)
21. Drewniak, L.; Styczek, A.; Majder-Lopatka, M.; Skłodowska, A. Bacteria, hypertolerant to arsenic in the rocks of an ancient gold mine, and their potential role in dissemination of arsenic pollution. *Environ. Pollut.* **2008**, *156*, 1069–1074. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Tomczyk-Żak, K.; Kaczanowski, S.; Drewniak, Ł.; Dmoch, Ł.; Skłodowska, A.; Zielenkiewicz, U. Bacteria diversity and arsenic mobilization in rock biofilm from an ancient gold and arsenic mine. *Sci. Total Environ.* **2013**, *461–462*, 330–340. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Drewniak, L.; Stasiuk, R.; Uhrynowski, W.; Skłodowska, A. *Shewanella* sp. O23S as a Driving Agent of a System Utilizing Dissimilatory Arsenate-Reducing Bacteria Responsible for Self-Cleaning of Water Contaminated with Arsenic. *Int. J. Mol. Sci.* **2015**, *16*, 14409–14427. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Romaniuk, K.; Dziewit, L.; Decewicz, P.; Mielnicki, S.; Radlinska, M.; Drewniak, L. Molecular characterization of the pSinB plasmid of the arsenite oxidizing, metalotolerant *Sinorhizobium* sp. M14—insight into the heavy metal resistome of sinorhizobial extrachromosomal replicons. *FEMS Microbiol. Ecol.* **2017**, *93*, fiw215. [\[CrossRef\]](#)
25. Uhrynowski, W.; Decewicz, P.; Dziewit, L.; Radlinska, M.; Krawczyk, P.S.; Lipinski, L.; Adamska, D.; Drewniak, L. Analysis of the Genome and Mobilome of a Dissimilatory Arsenate Reducing *Aeromonas* sp. O23A Reveals Multiple Mechanisms for Heavy Metal Resistance and Metabolism. *Front. Microbiol.* **2017**, *8*, 936. [\[CrossRef\]](#)
26. Whitman, W.B.; Coleman, D.C.; Wiebe, W.J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6578–6583. [\[CrossRef\]](#)
27. Naureen, Z.; Dautaj, A.; Anpilogov, K.; Camilleri, G.; Dhuli, K.; Tanzi, B.; Maltese, P.E.; Cristofoli, F.; De Antoni, L.; Beccari, T.; et al. Bacteriophages presence in nature and their role in the natural selection of bacterial populations. *Acta Biomed.* **2020**, *91*, e2020024. [\[CrossRef\]](#)
28. Hendrix, R.W. Bacteriophage genomics. *Curr. Opin. Microbiol.* **2003**, *6*, 506–511. [\[CrossRef\]](#)
29. Howard-Varona, C.; Hargreaves, K.R.; Abedon, S.T.; Sullivan, M.B. Lysogeny in nature: Mechanisms, impact and ecology of temperate phages. *ISME J.* **2017**, *11*, 1511–1520. [\[CrossRef\]](#)
30. Brüssow, H.; Hendrix, R.W. Phage genomics: Small is beautiful. *Cell* **2002**, *108*, 13–16. [\[CrossRef\]](#)
31. Canchaya, C.; Fournous, G.; Brüssow, H. The impact of prophages on bacterial chromosomes. *Mol. Microbiol.* **2004**, *53*, 9–18. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Fortier, L.C.; Sekulovic, O. Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence* **2013**, *4*, 354–365. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Jaroszewicz, W.; Morcinek-Orłowska, J.; Pierzynowska, K.; Gaffke, L.; Węgrzyn, G. Phage display and other peptide display technologies. *FEMS Microbiol. Rev.* **2022**, *46*, fuab052. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Latka, A.; Maciejewska, B.; Majkowska-Skrobek, G.; Briers, Y.; Drulis-Kawa, Z. Bacteriophage-encoded virion-associated enzymes to overcome the carbohydrate barriers during the infection process. *Appl. Microbiol. Biotechnol.* **2017**, *101*, 3103–3119. [\[CrossRef\]](#)
35. Orazi, G.; Collins, A.J.; Whitaker, R.J. Prediction of Prophages and Their Host Ranges in Pathogenic and Commensal. *mSystems* **2022**, *7*, e0008322. [\[CrossRef\]](#)
36. Sides, K.E. Agricultural Soil Bacteria; A Study of Collection, Cultivation, and Lysogeny. Master's Thesis, University of Tennessee, Knoxville, TN, USA, 2010.
37. Bahar, M.M.; Megharaj, M.; Naidu, R. Kinetics of arsenite oxidation by *Variovorax* sp. MM-1 isolated from a soil and identification of arsenite oxidase gene. *J. Hazard. Mater.* **2013**, *262*, 997–1003. [\[CrossRef\]](#)
38. Bachate, S.P.; Cavalca, L.; Andreoni, V. Arsenic-resistant bacteria isolated from agricultural soils of Bangladesh and characterization of arsenate-reducing strains. *J. Appl. Microbiol.* **2009**, *107*, 145–156. [\[CrossRef\]](#)



39. Akhter, S.; Aziz, R.K.; Edwards, R.A. PhiSpy: A novel algorithm for finding prophages in bacterial genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res.* **2012**, *40*, e126. [\[CrossRef\]](#)
40. Garneau, J.R.; Depardieu, F.; Fortier, L.C.; Bikard, D.; Monot, M. PhageTerm: A tool for fast and accurate determination of phage termini and packaging mechanism using next-generation sequencing data. *Sci. Rep.* **2017**, *7*, 8292. [\[CrossRef\]](#)
41. Recktenwald, J.; Schmidt, H. The nucleotide sequence of Shiga toxin (Stx) 2e-encoding phage phiP27 is not related to other Stx phage genomes, but the modular genetic structure is conserved. *Infect. Immun.* **2002**, *70*, 1896–1908. [\[CrossRef\]](#)
42. Feiss, M.; Rao, V.B. The bacteriophage DNA packaging machine. *Adv. Exp. Med. Biol.* **2012**, *726*, 489–509. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Kala, S.; Cumby, N.; Sadowski, P.D.; Hyder, B.Z.; Kanelis, V.; Davidson, A.R.; Maxwell, K.L. HNH proteins are a widespread component of phage DNA packaging machines. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 6022–6027. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Olszak, T.; Shneider, M.M.; Latka, A.; Maciejewska, B.; Browning, C.; Sycheva, L.V.; Cornelissen, A.; Danis-Wlodarczyk, K.; Senchenkova, S.N.; Shashkov, A.S.; et al. The O-specific polysaccharide lyase from the phage LKA1 tailspike reduces *Pseudomonas* virulence. *Sci. Rep.* **2017**, *7*, 16302. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Frick, D.N.; Richardson, C.C. DNA primases. *Annu. Rev. Biochem.* **2001**, *70*, 39–80. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Liu, Y.; Zheng, K.; Liu, B.; Liang, Y.; You, S.; Zhang, W.; Zhang, X.; Jie, Y.; Shao, H.; Jiang, Y.; et al. Characterization and Genomic Analysis of Marinobacter Phage vB\_MalS-PS3, Representing a New Lambda-Like Temperate Siphoviral Genus Infecting Algae-Associated Bacteria. *Front. Microbiol.* **2021**, *12*, 726074. [\[CrossRef\]](#)
47. Tisza, M.J.; Buck, C.B. A catalog of tens of thousands of viruses from human metagenomes reveals hidden associations with chronic diseases. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023202118. [\[CrossRef\]](#)
48. Roberts, J.W.; Yarnell, W.; Bartlett, E.; Guo, J.; Marr, M.; Ko, D.C.; Sun, H.; Roberts, C.W. Antitermination by bacteriophage lambda Q protein. *Cold Spring Harb. Symp. Quant. Biol.* **1998**, *63*, 319–325. [\[CrossRef\]](#)
49. Kropinski, A.M. Sequence of the genome of the temperate, serotype-converting, *Pseudomonas aeruginosa* bacteriophage D3. *J. Bacteriol.* **2000**, *182*, 6066–6074. [\[CrossRef\]](#)
50. Krylov, S.V.; Kropinski, A.M.; Shaburova, O.V.; Miroshnikov, K.A.; Chesnokova, E.N.; Krylov, V.N. New temperate *Pseudomonas aeruginosa* phage, phi297: Specific features of genome structure. *Genetika* **2013**, *49*, 930–942. [\[CrossRef\]](#)
51. Taylor, V.L.; Hoage, J.F.; Thrane, S.W.; Huszczyński, S.M.; Jelsbak, L.; Lam, J.S. A Bacteriophage-Acquired O-Antigen Polymerase (Wzy $\beta$ ) from *P. aeruginosa* Serotype O16 Performs a Varied Mechanism Compared to Its Cognate Wzy $\alpha$ . *Front. Microbiol.* **2016**, *7*, 393. [\[CrossRef\]](#)
52. Han, W.; Wu, B.; Li, L.; Zhao, G.; Woodward, R.; Pettit, N.; Cai, L.; Thon, V.; Wang, P.G. Defining function of lipopolysaccharide O-antigen ligase WaaL using chemoenzymatically synthesized substrates. *J. Biol. Chem.* **2012**, *287*, 5357–5365. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Newton, G.J.; Daniels, C.; Burrows, L.L.; Kropinski, A.M.; Clarke, A.J.; Lam, J.S. Three-component-mediated serotype conversion in *Pseudomonas aeruginosa* by bacteriophage D3. *Mol. Microbiol.* **2001**, *39*, 1237–1247. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Bin Jang, H.; Bolduc, B.; Zablocki, O.; Kuhn, J.H.; Roux, S.; Adriaenssens, E.M.; Brister, J.R.; Kropinski, A.M.; Krupovic, M.; Lavigne, R.; et al. Taxonomic assignment of uncultivated prokaryotic virus genomes is enabled by gene-sharing networks. *Nat. Biotechnol.* **2019**, *37*, 632–639. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Nishimura, Y.; Yoshida, T.; Kuronishi, M.; Uehara, H.; Ogata, H.; Goto, S. ViPTree: The viral proteomic tree server. *Bioinformatics* **2017**, *33*, 2379–2380. [\[CrossRef\]](#)
56. Trotereau, A.; Boyer, C.; Bornard, I.; Pécheur, M.J.B.; Schouler, C.; Torres-Barceló, C. High genomic diversity of novel phages infecting the plant pathogen *Ralstonia solanacearum*, isolated in Mauritius and Reunion islands. *Sci. Rep.* **2021**, *11*, 5382. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Yoon, S.H.; Ha, S.M.; Lim, J.; Kwon, S.; Chun, J. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie Van Leeuwenhoek* **2017**, *110*, 1281–1286. [\[CrossRef\]](#)
58. Adriaenssens, E.; Brister, J.R. How to Name and Classify Your Phage: An Informal Guide. *Viruses* **2017**, *9*, 70. [\[CrossRef\]](#)
59. Turner, D.; Kropinski, A.M.; Adriaenssens, E.M. A Roadmap for Genome-Based Phage Taxonomy. *Viruses* **2021**, *13*, 506. [\[CrossRef\]](#)
60. Decewicz, P.; Radlinska, M.; Dziewit, L. Characterization of *Sinorhizobium* sp. LM21 Prophages and Virus-Encoded DNA Methyltransferases in the Light of Comparative Genomic Analyses of the Sinorhizobial Virome. *Viruses* **2017**, *9*, 161. [\[CrossRef\]](#)
61. Decewicz, P.; Dziewit, L.; Golec, P.; Kozłowska, P.; Bartosik, D.; Radlinska, M. Characterization of the virome of *Paracoccus* spp. (Alphaproteobacteria) by combined in silico and in vivo approaches. *Sci. Rep.* **2019**, *9*, 7899. [\[CrossRef\]](#)
62. Altschul, S.F.; Madden, T.L.; Schäffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [\[CrossRef\]](#) [\[PubMed\]](#)
63. Dziurzynski, M.; Decewicz, P.; Ciuchcinski, K.; Gorecki, A.; Dziewit, L. Simple, Reliable, and Time-Efficient Manual Annotation of Bacterial Genomes with MAISEN. *Methods Mol. Biol.* **2021**, *2242*, 221–229. [\[CrossRef\]](#) [\[PubMed\]](#)
64. Williams, K.P. Integration sites for genetic elements in prokaryotic tRNA and tmRNA genes: Sublocation preference of integrase subfamilies. *Nucleic Acids Res.* **2002**, *30*, 866–875. [\[CrossRef\]](#) [\[PubMed\]](#)
65. Canchaya, C.; Proux, C.; Fournous, G.; Bruttin, A.; Brüssow, H. Prophage genomics. *Microbiol. Mol. Biol. Rev.* **2003**, *67*, 238–276. [\[CrossRef\]](#)
66. Biosca, E.G.; Català-Senent, J.F.; Figàs-Segura, À.; Bertolini, E.; López, M.M.; Álvarez, B. Genomic Analysis of the First European Bacteriophages with Depolymerase Activity and Biocontrol Efficacy against the Phytopathogen. *Viruses* **2021**, *13*, 2539. [\[CrossRef\]](#) [\[PubMed\]](#)



67. Santamaría, R.I.; Bustos, P.; Sepúlveda-Robles, O.; Lozano, L.; Rodríguez, C.; Fernández, J.L.; Juárez, S.; Kameyama, L.; Guarneros, G.; Dávila, G.; et al. Narrow-host-range bacteriophages that infect *Rhizobium etli* associate with distinct genomic types. *Appl. Environ. Microbiol.* **2014**, *80*, 446–454. [\[CrossRef\]](#)
68. Ely, B.; Berrios, L.; Thomas, Q. S2B, a Temperate Bacteriophage That Infects *Caulobacter Crescentus* Strain CB15. *Curr. Microbiol.* **2022**, *79*, 98. [\[CrossRef\]](#)
69. Sambrook, J.; Russell, D.W. *Molecular Cloning: A Laboratory Manual*, 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, USA, 2001.
70. Carbajal-Rodríguez, I.; Stöveken, N.; Satola, B.; Wübbeler, J.H.; Steinbüchel, A. Aerobic degradation of mercaptosuccinate by the gram-negative bacterium *Variovorax paradoxus* strain B4. *J. Bacteriol.* **2011**, *193*, 527–539. [\[CrossRef\]](#)
71. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **2018**, *7*, 1338. [\[CrossRef\]](#)
72. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [\[CrossRef\]](#)
73. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:20131303.3997. [\[CrossRef\]](#)
75. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R.; Subgroup, G.P.D.P. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [\[CrossRef\]](#) [\[PubMed\]](#)
76. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [\[CrossRef\]](#)
77. Bujak, K.; Decewicz, P.; Kaminski, J.; Radlinska, M. Identification, Characterization, and Genomic Analysis of Novel. *Int. J. Mol. Sci.* **2020**, *21*, 6709. [\[CrossRef\]](#)
78. Carver, T.; Berriman, M.; Tivey, A.; Patel, C.; Böhme, U.; Barrell, B.G.; Parkhill, J.; Rajandream, M.A. Artemis and ACT: Viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* **2008**, *24*, 2672–2676. [\[CrossRef\]](#)
79. Brettin, T.; Davis, J.J.; Disz, T.; Edwards, R.A.; Gerdes, S.; Olsen, G.J.; Olson, R.; Overbeek, R.; Parrello, B.; Pusch, G.D.; et al. RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **2015**, *5*, 8365. [\[CrossRef\]](#)
80. Wattam, A.R.; Davis, J.J.; Assaf, R.; Boisvert, S.; Brettin, T.; Bun, C.; Conrad, N.; Dietrich, E.M.; Disz, T.; Gabbard, J.L.; et al. Improvements to PATRIC, the all-bacterial Bioinformatics Database and Analysis Resource Center. *Nucleic Acids Res.* **2017**, *45*, D535–D542. [\[CrossRef\]](#)
81. Hildebrand, A.; Remmert, M.; Biegert, A.; Söding, J. Fast and accurate automatic structure prediction with HHpred. *Proteins* **2009**, *77* (Suppl. S9), 128–132. [\[CrossRef\]](#)
82. Blum, M.; Chang, H.Y.; Chuguransky, S.; Grego, T.; Kandasaamy, S.; Mitchell, A.; Nuka, G.; Paysan-Lafosse, T.; Qureshi, M.; Raj, S.; et al. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res.* **2021**, *49*, D344–D354. [\[CrossRef\]](#)
83. Finn, R.D.; Bateman, A.; Clements, J.; Coghill, P.; Eberhardt, R.Y.; Eddy, S.R.; Heger, A.; Hetherington, K.; Holm, L.; Mistry, J.; et al. Pfam: The protein families database. *Nucleic Acids Res.* **2014**, *42*, D222–D230. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Apweiler, R.; Bairoch, A.; Wu, C.H.; Barker, W.C.; Boeckmann, B.; Ferro, S.; Gasteiger, E.; Huang, H.; Lopez, R.; Magrane, M.; et al. UniProt: The Universal Protein knowledgebase. *Nucleic Acids Res.* **2004**, *32*, D115–D119. [\[CrossRef\]](#) [\[PubMed\]](#)
85. Krogh, A.; Larsson, B.; von Heijne, G.; Sonnhammer, E.L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **2001**, *305*, 567–580. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Schattner, P.; Brooks, A.N.; Lowe, T.M. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **2005**, *33*, W686–W689. [\[CrossRef\]](#)
87. Laslett, D.; Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **2004**, *32*, 11–16. [\[CrossRef\]](#)
88. Lopes, A.; Tavares, P.; Petit, M.A.; Guérois, R.; Zinn-Justin, S. Automated classification of tailed bacteriophages according to their neck organization. *BMC Genomics* **2014**, *15*, 1027. [\[CrossRef\]](#)
89. Gilchrist, C.L.M.; Chooi, Y.H. Clinker & clustermap.js: Automatic generation of gene cluster comparison figures. *Bioinformatics* **2021**, *37*, 2473–2475. [\[CrossRef\]](#)
90. Cook, R.; Brown, N.; Redgwell, T.; Rihtman, B.; Barnes, M.; Clokie, M.; Stekel, D.J.; Hobman, J.; Jones, M.A.; Millard, A. INfrastructure for a PHAge REference Database: Identification of Large-Scale Biases in the Current Collection of Cultured Phage Genomes. *PHAGE* **2021**, *2*, 214–223. [\[CrossRef\]](#)
91. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the International AAAI Conference on Web and Social Media, San Jose, CA, USA, 17–20 May 2009; Volume 3, pp. 361–362. Available online: <https://ojs.aaai.org/index.php/ICWSM/article/view/13937> (accessed on 5 October 2022). [\[CrossRef\]](#)
92. Jacomy, M.; Venturini, T.; Heymann, S.; Bastian, M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS ONE* **2014**, *9*, e98679. [\[CrossRef\]](#)
93. Jackson, C.R.; Harrison, K.G.; Dugas, S.L. Enumeration and characterization of culturable arsenate resistant bacteria in a large estuary. *Syst. Appl. Microbiol.* **2005**, *28*, 727–734. [\[CrossRef\]](#) [\[PubMed\]](#)

- 
94. Flores-Duarte, N.J.; Pérez-Pérez, J.; Navarro-Torre, S.; Mateos-Naranjo, E.; Redondo-Gómez, S.; Pajuelo, E.; Rodríguez-Llorente, I.D. Improved. *Plants* **2022**, *11*, 1091. [[CrossRef](#)] [[PubMed](#)]
  95. Kämpfer, P.; Busse, H.J.; McInroy, J.A.; Glaeser, S.P. *Variovorax gossypii* sp. nov., isolated from *Gossypium hirsutum*. *Int. J. Syst. Evol. Microbiol.* **2015**, *65*, 4335–4340. [[CrossRef](#)] [[PubMed](#)]
  96. Han, J.I.; Choi, H.K.; Lee, S.W.; Orwin, P.M.; Kim, J.; Laroe, S.L.; Kim, T.G.; O’Neil, J.; Leadbetter, J.R.; Lee, S.Y.; et al. Complete genome sequence of the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. *J. Bacteriol.* **2011**, *193*, 1183–1190. [[CrossRef](#)] [[PubMed](#)]